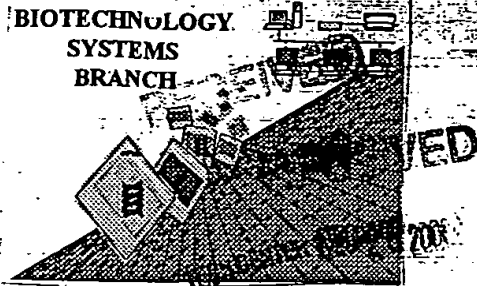


*UP*

## RAW SEQUENCE LISTING ERROR REPORT



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer-readable form:

Application Serial Number: 09/424,482

Source: 1627

Date Processed by STIC: 12/11/2000

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: [patin30help@uspto.gov](mailto:patin30help@uspto.gov) or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

### Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 - 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO).

Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:

<http://www.uspto.gov/web/offices/pac/checker>

# Raw Sequence Listing Error Summary

## ERROR DETECTED SUGGESTED CORRECTION

SERIAL NUMBER: 09/424,482

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1 ☐ **Wrapped Nucleics** The number/text at the end of each line "wrapped" down to the next line.  
This may occur if your file was retrieved in a word processor after creating it.  
Please adjust your right margin to .3, as this will prevent "wrapping".
- 2 ☐ **Wrapped Aminos** The amino acid number/text at the end of each line "wrapped" down to the next line.  
This may occur if your file was retrieved in a word processor after creating it.  
Please adjust your right margin to .3, as this will prevent "wrapping".
- 3 ☐ **Incorrect Line Length** The rules require that a line not exceed 72 characters in length. This includes spaces.
- 4 ☐ **Misaligned Amino Acid Numbering** The numbering under each 5th amino acid is misaligned. This may be caused by the use of tabs between the numbering. It is recommended to delete any tabs and use spacing between the numbers.
- 5 ☐ **Non-ASCII** This file was not saved in ASCII (DOS) text, as required by the Sequence Rules.  
Please ensure your subsequent submission is saved in ASCII text so that it can be processed.
- 6 ☒ **Variable Length** Sequence(s) 4-5 (maybe more) contain n's or Xaa's which represented more than one residue.  
As per the rules, each n or Xaa can only represent a single residue.  
Please present the maximum number of each residue having variable length and indicate in the (ix) feature section that some may be missing.
- 7 ☐ **PatentIn ver. 2.0 "bug"** A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequence(s) \_\_\_\_\_. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies primarily to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
- 8 ☐ **Skipped Sequences (OLD RULES)** Sequence(s) \_\_\_\_ missing. If intentional, please use the following format for each skipped sequence:  
(2) INFORMATION FOR SEQ ID NO:X:  
(i) SEQUENCE CHARACTERISTICS: (Do not insert any headings under "SEQUENCE CHARACTERISTICS")  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:X:  
This sequence is intentionally skipped  
  
Please also adjust the "(iii) NUMBER OF SEQUENCES:" response to include the skipped sequence(s).
- 9 ☐ **Skipped Sequences (NEW RULES)** Sequence(s) \_\_\_\_ missing. If intentional, please use the following format for each skipped sequence.  
<210> sequence id number  
<400> sequence id number  
000
- 10 ☐ **Use of n's or Xaa's (NEW RULES)** Use of n's and/or Xaa's have been detected in the Sequence Listing.  
Use of <220> to <223> is MANDATORY if n's or Xaa's are present.  
In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
- 11 ☐ **Use of <213>Organism (NEW RULES)** Sequence(s) \_\_\_\_\_ are missing this mandatory field or its response.
- 12 ☐ **Use of <220>Feature (NEW RULES)** Sequence(s) \_\_\_\_ are missing the <220>Feature and associated headings.  
Use of <220> to <223> is MANDATORY if <213>ORGANISM is "Artificial" or "Unknown"  
Please explain source of genetic material in <220> to <223> section.  
(See "Federal Register," 6/01/98, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of new Rules)
- 13 ☐ **PatentIn ver. 2.0 "bug"** Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing).  
Instead, please use "File Manager" or any other means to copy file to floppy disk.

1627

Does Not Comply  
Corrected Diskette NeededRAW SEQUENCE LISTING  
PATENT APPLICATION: US/09/424,482DATE: 12/11/2000  
TIME: 16:39:46Input Set : A:\CHOO 424482 SEQUENCE.txt  
Output Set: N:\CRF3\12112000\I424482.raw

PP-5,2-3

3 <110> APPLICANT: Medical Research Council  
4 Choo, Yen  
5 Klug, Aaron  
6 Isalan, Mark  
8 <120> TITLE OF INVENTION: Nucleic Acid Binding Polypeptide Library  
10 <130> FILE REFERENCE: 71278/264974  
12 <140> CURRENT APPLICATION NUMBER: US 09/424,482  
C--> 13 <141> CURRENT FILING DATE: 2000-02-29  
15 <150> PRIOR APPLICATION NUMBER: GB9710809.6  
16 <151> PRIOR FILING DATE: 1997-05-23  
18 <150> PRIOR APPLICATION NUMBER: PCT/GB98/01510  
19 <151> PRIOR FILING DATE: 1998-05-25  
21 <160> NUMBER OF SEQ ID NOS: 19  
23 <170> SOFTWARE: PatentIn version 3.0  
25 <210> SEQ ID NO: 1  
26 <211> LENGTH: 9  
27 <212> TYPE: DNA  
28 <213> ORGANISM: Artificial  
30 <220> FEATURE:  
31 <223> OTHER INFORMATION: Description of Artificial Sequence: LIB-A DNA sorting sequence  
33 <220> FEATURE:  
34 <221> NAME/KEY: variation  
35 <222> LOCATION: (2)..(4)  
36 <223> OTHER INFORMATION: n is any nucleotide  
39 <400> SEQUENCE: 1  
OK 40 gnnncggcg 9  
43 <210> SEQ ID NO: 2  
44 <211> LENGTH: 9  
45 <212> TYPE: DNA  
46 <213> ORGANISM: Artificial  
48 <220> FEATURE:  
49 <223> OTHER INFORMATION: Description of Artificial Sequence: LIB-B DNA sorting sequence  
51 <220> FEATURE:  
52 <221> NAME/KEY: variation  
53 <222> LOCATION: (3)..(4)  
54 <223> OTHER INFORMATION: n is any other nucleotide  
OK 57 <400> SEQUENCE: 2 9  
58 gcnnccggcg  
61 <210> SEQ ID NO: 3  
62 <211> LENGTH: 9  
63 <212> TYPE: DNA  
64 <213> ORGANISM: Artificial  
66 <220> FEATURE:  
67 <223> OTHER INFORMATION: Description of Artificial DNA: LIB 1/2 sorting sequence  
69 <220> FEATURE:  
70 <221> NAME/KEY: variation  
71 <222> LOCATION: (5)..(9)

## RAW SEQUENCE LISTING

DATE: 12/11/2000

PATENT APPLICATION: US/09/424,482

TIME: 16:39:46

Input Set : A:\CHOO 424482 SEQUENCE.txt

Output Set: N:\CRF3\12112000\1424482.raw

72 <223> OTHER INFORMATION: n is any other nucleotide  
 75 <400> SEQUENCE: 3  
 76 gcggnnnnn  
 79 <210> SEQ ID NO: 4  
 80 <211> LENGTH: 18  
 81 <212> TYPE: PRT  
 82 <213> ORGANISM: Artificial  
 84 <220> FEATURE:  
 85 <223> OTHER INFORMATION: Description of Artificial Sequence: Structure A  
 87 <220> FEATURE:  
 88 <221> NAME/KEY: SITE  
 89 <222> LOCATION: (1)..(1)  
 90 <223> OTHER INFORMATION: Xaa is any amino acid  
 93 <220> FEATURE:  
 94 <221> NAME/KEY: SITE  
 95 <222> LOCATION: (3)..(3)  
 96 <223> OTHER INFORMATION: Xaa is any amino acid  
 99 <220> FEATURE:  
 100 <221> NAME/KEY: SITE  
 101 <222> LOCATION: (5)..(13)  
 102 <223> OTHER INFORMATION: Xaa is any amino acid  
 105 <220> FEATURE:  
 106 <221> NAME/KEY: SITE  
 107 <222> LOCATION: (15)..(17)  
 108 <223> OTHER INFORMATION: Xaa is any amino acid  
 111 <220> FEATURE:  
 112 <221> NAME/KEY: VARIANT  
 113 <222> LOCATION: (1)..(1)  
 114 <223> OTHER INFORMATION: 0 - 2 possible residues  
 117 <220> FEATURE:  
 118 <221> NAME/KEY: VARIANT  
 119 <222> LOCATION: (3)..(3)  
 120 <223> OTHER INFORMATION: 1 - 5 possible residues  
 123 <220> FEATURE:  
 124 <221> NAME/KEY: VARIANT  
 125 <222> LOCATION: (5)..(13)  
 126 <223> OTHER INFORMATION: 9 - 14 possible residues  
 129 <220> FEATURE:  
 130 <221> NAME/KEY: VARIANT  
 131 <222> LOCATION: (15)..(17)  
 132 <223> OTHER INFORMATION: 3 - 6 possible residues  
 135 <220> FEATURE:  
 136 <221> NAME/KEY: SITE  
 137 <222> LOCATION: (18)..(18)  
 138 <223> OTHER INFORMATION: X is His or Cys  
 141 <400> SEQUENCE: 7  
 W--> 143 Xaa Cys Xaa Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His Xaa Xaa  
 144 5 10 15  
 W--> 146 Xaa Xaa

variable length is  
 not permitted  
 (see item 6 on  
 Erra Summary  
 Sheet)

RAW SEQUENCE LISTING                      DATE: 12/11/2000  
 PATENT APPLICATION: US/09/424,482      TIME: 16:39:46

Input Set : A:\CH00 424482 SEQUENCE.txt  
 Output Set: N:\CRF3\12112000\I424482.raw

```

149 <210> SEQ ID NO: 5
150 <211> LENGTH: 21
151 <212> TYPE: PRT
152 <213> ORGANISM: Artificial
154 <220> FEATURE:
155 <223> OTHER INFORMATION: Description of Artificial Sequence: Structure B
157 <220> FEATURE:
158 <221> NAME/KEY: SITE
159 <222> LOCATION: (1)..(1)
160 <223> OTHER INFORMATION: Xaa is any amino acid
163 <220> FEATURE:
164 <221> NAME/KEY: SITE
165 <222> LOCATION: (3)..(4)
166 <223> OTHER INFORMATION: Xaa is any amino acid
169 <220> FEATURE:
170 <221> NAME/KEY: SITE
171 <222> LOCATION: (6)..(7)
172 <223> OTHER INFORMATION: Xaa is any amino acid
175 <220> FEATURE:
176 <221> NAME/KEY: SITE
177 <222> LOCATION: (9)..(13)
178 <223> OTHER INFORMATION: Xaa is any amino acid
181 <220> FEATURE:
182 <221> NAME/KEY: SITE
183 <222> LOCATION: (15)..(16)
184 <223> OTHER INFORMATION: Xaa is any amino acid
187 <220> FEATURE:
188 <221> NAME/KEY: SITE
189 <222> LOCATION: (18)..(20)
190 <223> OTHER INFORMATION: Xaa is any amino acid
193 <220> FEATURE:
194 <221> NAME/KEY: VARIANT
195 <222> LOCATION: (3)..(4)
196 <223> OTHER INFORMATION: 2 or 4 amino acids
199 <220> FEATURE:
200 <221> NAME/KEY: VARIANT
201 <222> LOCATION: (6)..(7)
202 <223> OTHER INFORMATION: 2 or 3 amino acids
205 <400> SEQUENCE: 5
W--> 207 Xaa Cys Xaa Xaa Cys Xaa Xaa Phe Xaa Xaa Xaa Xaa Leu Xaa Xaa
      208 1      5      10      15
W--> 210 His Xaa Xaa Xaa His
      211      20
213 <210> SEQ ID NO: 6
214 <211> LENGTH: 4
215 <212> TYPE: PRT
216 <213> ORGANISM: Artificial
218 <220> FEATURE:
219 <223> OTHER INFORMATION: Description of Artificial Sequence: Linker

```

*variable length is not permitted  
 (see item 6 on  
 Error Summary  
 sheet)*

RAW SEQUENCE LISTING                      DATE: 12/11/2000  
 PATENT APPLICATION: US/09/424,482        TIME: 16:39:46

Input Set : A:\CH00 424482 SEQUENCE.txt  
 Output Set: N:\CRF3\12112000\I424482.raw

```

221 <400> SEQUENCE: 6
223 Thr Gly Glu Lys
224 1
226 <210> SEQ ID NO: 7
227 <211> LENGTH: 5
228 <212> TYPE: PRT
229 <213> ORGANISM: Artificial
231 <220> FEATURE:
232 <223> OTHER INFORMATION: Description of Artificial Sequence: Linker
234 <400> SEQUENCE: 7
236 Thr Gly Glu Lys Pro
237 1 5
239 <210> SEQ ID NO: 8
240 <211> LENGTH: 26
241 <212> TYPE: PRT
242 <213> ORGANISM: Artificial
244 <220> FEATURE:
245 <223> OTHER INFORMATION: Description of Artificial Sequence: Consensus structure
247 <400> SEQUENCE: 8
249 Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Lys Ser Asp
250 1 5 10 15
252 Leu Val Lys His Gln Arg Thr His Thr Gly
253 20 25
255 <210> SEQ ID NO: 9
256 <211> LENGTH: 29
257 <212> TYPE: PRT
258 <213> ORGANISM: Artificial
260 <220> FEATURE:
261 <223> OTHER INFORMATION: Description of Artificial Sequence: Consensus structure
263 <400> SEQUENCE: 9
265 Pro Tyr Lys Cys Ser Glu Cys Gly Lys Ala Phe Ser Gln Lys Ser Asn
266 1 5 10 15
268 Leu Thr Arg His Gln Arg Ile His Thr Gly Glu Lys Pro
269 20 25
271 <210> SEQ ID NO: 10
272 <211> LENGTH: 6
273 <212> TYPE: PRT
274 <213> ORGANISM: Artificial
276 <220> FEATURE:
277 <223> OTHER INFORMATION: Description of Artificial Sequence: Leader peptide
279 <400> SEQUENCE: 10
281 Met Ala Glu Glu Lys Pro
282 1 5
284 <210> SEQ ID NO: 11
285 <211> LENGTH: 9
286 <212> TYPE: DNA
287 <213> ORGANISM: Artificial
289 <220> FEATURE:
290 <223> OTHER INFORMATION: Description of Artificial DNA: LIB 2/3 DNA sorting sequence

```

RAW SEQUENCE LISTING      DATE: 12/11/2000  
 PATENT APPLICATION: US/09/424,482      TIME: 16:39:46

Input Set : A:\CH00 424482 SEQUENCE.txt  
 Output Set: N:\CRF3\12112000\I424482.raw

RECEIVED

DEC 26 2000

TECH CENTER 1600/2900

292 <220> FEATURE:  
 293 <221> NAME/KEY: variation  
 294 <222> LOCATION: (1)..(5)  
 295 <223> OTHER INFORMATION: n is any nucleotide  
 298 <400> SEQUENCE: 11  
 OK> 299 nnnnnnggcg  
 302 <210> SEQ ID NO: 12  
 303 <211> LENGTH: 9  
 304 <212> TYPE: DNA  
 305 <213> ORGANISM: Artificial  
 307 <220> FEATURE:  
 308 <223> OTHER INFORMATION: Description of Artificial Sequence: Zinc finger -DNA interaction  
 309      sequence      Sequence  
 311 <400> SEQUENCE: 12  
 312 cgcgccacgc      9  
 315 <210> SEQ ID NO: 13  
 316 <211> LENGTH: 9  
 317 <212> TYPE: DNA  
 318 <213> ORGANISM: Artificial  
 320 <220> FEATURE:  
 321 <223> OTHER INFORMATION: Description of Artificial Sequence: Zinc finger-DNA interaction s  
 322      sequence      Sequence      ← move down  
 324 <400> SEQUENCE: 13  
 325 acgcgccacg      9  
 328 <210> SEQ ID NO: 14  
 329 <211> LENGTH: 9  
 330 <212> TYPE: DNA  
 331 <213> ORGANISM: Artificial  
 333 <220> FEATURE:  
 334 <223> OTHER INFORMATION: Description of Artificial Sequence: Zinc finger-DNA interaction s  
 335      sequence      Sequence      ←  
 337 <400> SEQUENCE: 14  
 338 gcgtgggcg      9  
 341 <210> SEQ ID NO: 15  
 342 <211> LENGTH: 9  
 343 <212> TYPE: DNA  
 344 <213> ORGANISM: Artificial  
 346 <220> FEATURE:  
 347 <223> OTHER INFORMATION: Description of Artificial Sequence: Zinc finger-DNA interaction l  
 348      library designed sequence      ←  
 350 <220> FEATURE:  
 351 <221> NAME/KEY: variation  
 352 <222> LOCATION: (7)..(9)  
 353 <223> OTHER INFORMATION: n is any nucleotide  
 OK> 356 <400> SEQUENCE: 15  
 357 acgccgnnn      9  
 360 <210> SEQ ID NO: 16  
 361 <211> LENGTH: 36  
 362 <212> TYPE: PRT

FYI

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

VERIFICATION SUMMARY                      DATE: 12/11/2000  
PATENT APPLICATION: US/09/424,482        TIME: 16:39:47

Input Set : A:\CHOO 424482 SEQUENCE.txt  
Output Set: N:\CRF3\12112000\I424482.raw

L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:40 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:58 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2  
L:76 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:143 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4  
L:146 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4  
L:207 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5  
L:210 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5  
L:299 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11  
L:357 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:15  
L:400 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:17  
L:425 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:18  
L:447 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:19